

動的学習による辞書を用いたMatching Pursuits符号化

著者	今村 幸祐, 田巻 渉, 橋本 秀雄
雑誌名	映像情報メディア学会誌 : 映像情報メディア = Journal of the Institute of Image Information and Television Engineers
巻	63
号	1
ページ	105-109
発行年	2009-01-01
URL	http://hdl.handle.net/2297/17358

動的学習による辞書を用いた Matching Pursuits 符号化

Matching Pursuits Coding Using Dynamic Learning Dictionary

今村 幸祐[†], 田巻 渉[†], 正会員 橋本 秀雄[†]Kousuke Imamura[†], Wataru Tamaki[†] and Hideo Hashimoto[†]

Abstract Recently, an efficient video coding method at low bit rate using Matching Pursuits (MP) has been proposed. The MP coding method represents a signal in an approximate form using a dictionary. Therefore, coding performance depends greatly on the dictionary. In this paper, we introduce a video coding method that employs motion compensation and MP using a dynamic learning dictionary. The dictionary of the proposed method is renewed at each frame by using encoded information. Simulation results show that the coding performance of MP can be improved by applying the dynamic learning dictionary.

キーワード : video coding, Matching Pursuits, dynamic learning dictionary

1. ま え が き

動画像符号化技術は、地上デジタル放送や DVD, 携帯電話など、幅広いアプリケーションへ適用されている。動画像符号化技術の中心である MPEG や H.26x などの国際標準方式においては、時間的冗長度の削減のための動き補償予測 (MC) と空間的冗長度の削減のための離散コサイン変換 (DCT) を組合せたハイブリッド方式が基本である。それに対し近年では、新しい符号化方式の開発のため、DCT の代わりに他の方式を適用した動画像符号化方式の検討も行われている。その一つとして、Matching Pursuits (MP) と呼ばれる波形符号化法を用いた方式が、低ビットレートで高い符号化特性を得られる手法として報告されている¹⁾²⁾。

Matching Pursuits は、辞書に登録された基底ベクトルの線形結合によって、符号化対象信号を近似する手法である³⁾。したがって、Matching Pursuits の符号化特性は符号化に用いる辞書に大きく依存する。そのため、辞書の改善に対する検討も多く行われている。

辞書の改善のためのアプローチは、大きく二つに分けられる。一つは辞書構造を改善することにより、符号化特性と処理時間を改善するものである。これらには、階層的な辞書を用いることで符号化処理を高速化したものや⁴⁾、対象信号をサブバンド分割し、帯域毎に異なる辞書を用いる

ことにより、特性改善と高速化を実現する手法などが提案されている⁵⁾⁶⁾。

もう一つは、辞書に用いる基底ベクトルを、符号化対象画像に対して最適化するというアプローチである。本論文は、この辞書の最適化を中心に検討を行う。

辞書の最適化法としては、経験的な手法の他に、代表的なものとして、学習による辞書パラメータの最適化法が提案されている⁷⁾⁸⁾。これらの学習的手法は、符号化対象が既知の場合は、あらかじめ学習により最適化された辞書を作成することができる。しかし、一般的には符号化対象シーケンスは既知ではないため、通常は複数の学習シーケンスを用いて最適化を行い、準最適な汎用辞書を作成して適用することになる。つまり、これらの手法では、個々のシーケンスに適した辞書設計は困難である。また、シーケンス全体を通じて同一の辞書を用いるため、途中で特性が変化するシーケンスには充分に対応できないという問題もある。

本論文では、これらの問題の解決策として、符号化時に得られた情報により、動的な学習を行うことで、辞書を最適化する手法を提案する。この動的学習辞書による Matching Pursuits 符号化法を用いることで、既知ではない符号化対象シーケンスについても、最適な辞書を用いた高能率な符号化が可能となる。また本提案法では、各フレームの符号化情報を用いて、フレーム単位で辞書更新を行うため、途中で特性が変化するシーケンスについても対応が可能となる。

2. Matching Pursuits

2.1 Matching Pursuits の基本原理

Matching Pursuits は、非直交で overcomplete な基底ベクトル $g_j(t)$ によって構成される辞書 \mathcal{D} 内の波形の線

2008 年 7 月 30 日受付, 2008 年 10 月 10 日再受付, 2008 年 10 月 28 日採録

[†] 金沢大学 大学院 自然科学研究科

(〒 920-1192 石川県金沢市角間町, 076-234-4894)

[†] Graduate School of Natural Science & Technology, Kanazawa University

(Kakumamati, Kanazawa-shi, Ishikawa, 920-1192, Japan)

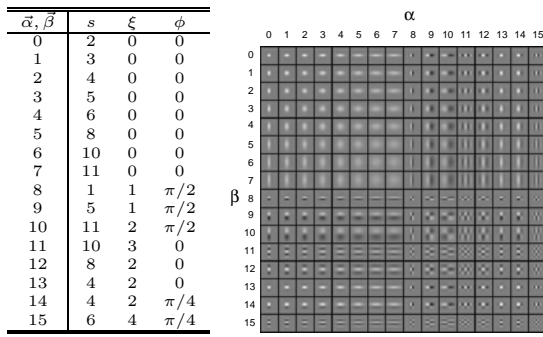


図 1 Matching Pursuits 基本辞書 (N=16)
Basic dictionary for Matching Pursuits (N=16).

形結合により, 符号化対象信号 $f(t)$ を近似構成する.

Matching Pursuits による近似信号 $\hat{f}(t)$ は式 (1) で表される.

$$\hat{f}(t) = \sum_{k=1}^m p_k \cdot g_{\gamma_k}(t - \tau_k) \quad (1)$$

ここで τ_k, γ_k, p_k はそれぞれ基底ベクトルの位置, 種類, スケールを表し, これらの 3 種類のパラメータの組 (τ_k, γ_k, p_k) によって決定する波形をアトム (atom) と呼ぶ. m は信号を構成するアトムの数を表す.

まず, 信号 $f(t)$ をある基底ベクトル $g_{\gamma_1}(t - \tau_1)$ へ射影することにより信号 $f(t)$ は

$$f(t) = p_1 \cdot g_{\gamma_1}(t - \tau_1) + Rf(t) \quad (2)$$

$$p_1 = \langle f(t), g_{\gamma_1}(t - \tau_1) \rangle \quad (3)$$

と表せる. ここで p_1 は信号と基底ベクトルの内積値であり, $Rf(t)$ は基底ベクトル $g_{\gamma_1}(t - \tau_1)$ へ射影した場合の残差信号である. また基底ベクトル $g_{\gamma}(t)$ のノルムが 1 に正規化されているものとする, 式 (2) において $Rf(t)$ と $g_{\gamma_1}(t - \tau_1)$ は互いに直交していることから, 以下の関係式が成立する.

$$\|f(t)\|^2 = p_1^2 + \|Rf(t)\|^2 \quad (4)$$

すなわち内積値 p_1 の絶対値を最大とするパラメータを求めることにより, 反復の各段階において残差信号のエネルギー減少量 $\Delta e = \|f(t)\|^2 - \|Rf(t)\|^2$ を最大とするアトムを決定することができる.

同様に残差信号 $Rf(t)$ を展開すると

$$Rf(t) = p_2 \cdot g_{\gamma_2}(t - \tau_2) + R^2 f(t) \quad (5)$$

となる. 上記の処理を繰り返すことにより, m 個のアトムにより展開された信号 $f(t)$ は

$$f(t) = \sum_{k=1}^m p_k \cdot g_{\gamma_k}(t - \tau_k) + R^m f(t) \quad (6)$$

となる. したがって, m 個のアトムにより信号 $f(t)$ を近似的に再構成することができる.

2.2 Matching Pursuits の辞書

Matching Pursuits では, 式 (7) に示す Gabor 関数の直積により作成される 2 次元可分型辞書が一般的に用いられる.

$$g_{\vec{\gamma}}(n) = K_{\vec{\gamma}} g\left(\frac{n - \frac{N}{2} + 1}{s}\right) \cos\left(\frac{2\pi\xi(n - \frac{N}{2} + 1)}{N} + \phi\right) \quad (7)$$

$$n \in \{0, 1, \dots, N - 1\}$$

ここで, K は正規化係数であり, N は基底関数長を示す. また, $g(\cdot)$ はガウシアンウィンドウを示す. 式 (7) から得られる波形に対して, s, ξ, ϕ は, それぞれガウシアンウィンドウのスケール, 変調周波数, 位相シフト量を決定するパラメータである.

さらに水平および垂直方向のための二つのパラメータセット $\bar{\alpha}, \bar{\beta}$ を用いて得られたベクトルと式 (8) により, 2 次元の可分型辞書を得ることができる.

$$G(i, j) = g_{\bar{\alpha}(i)} \otimes g_{\bar{\beta}(j)}, \quad \otimes: \text{直積} \quad (8)$$

図 1 に, 文献 2) で用いられているパラメータセット $\bar{\alpha}, \bar{\beta}$ と式 (7), (8) から作成される辞書を示す. $\bar{\alpha}, \bar{\beta}$ はそれぞれ辞書における水平および垂直方向の基底ベクトルを決定するパラメータセットである. また図 1 右側は, 実際に生成した 2 次元辞書波形を画像で表したものであり, 一つのブロックが水平・垂直それぞれ一つのパラメータセットに対応している.

2.3 静的学習による辞書設計

学習的手法を用いることで, 学習用シーケンスに対して, 辞書パラメータの最適化を行うことができる⁷⁾⁸⁾. このような手法は, あらかじめ決められた学習用シーケンスを用いて学習し, 符号化中に辞書が変化しないことから, 本論文では静的学習辞書と呼ぶものとする.

以下に基本的な静的学習辞書設計アルゴリズムを示す⁸⁾.

1. 学習用シーケンス全体に対して, Matching Pursuits を実行し, 学習情報として適用した辞書パラメータセットのインデックス i, j と, 符号化対象となったブロック X の画素値情報を取得する.

2. 辞書により近似される信号エネルギー E を, 辞書と符号化対象ブロックの内積の絶対値和である式 (9) で評価する.

$$E = \sum_{k=1}^{m_x} |\langle X_k, G(i_k, j_k) \rangle| \quad (9)$$

ここで m_x は, 学習情報を取得したブロック数を示す. 更新は, パラメータセット $\bar{\alpha}, \bar{\beta}$ における s, ξ, ϕ について, 設定した範囲内の離散値に対して網羅的に評価し, 評価式 E が最大となるパラメータを更新値として選択する.

3. 収束または最大繰り返し回数に達するまで 1 と 2 の処理を繰り返す.

この学習アルゴリズムの繰り返しにおける特性として,

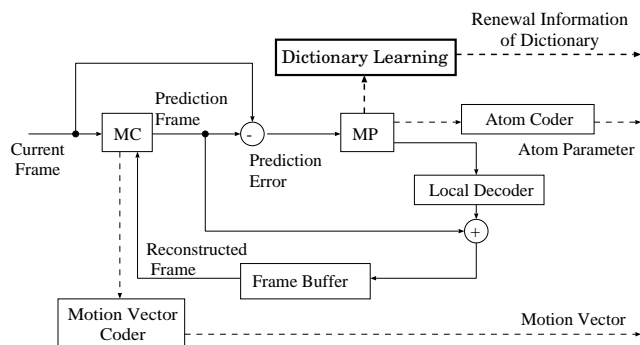


図2 動的辞書を用いた動画画像符号化システム
Video coding system using dynamic dictionary.

最初の処理において大きな特性改善があり、その後の繰り返し処理では改善量が少ないことが報告されている⁷⁾。

3. 動的学習による辞書設計

まず、動的学習辞書を用いた Matching Pursuits を適用した動画画像符号化方式について述べる。図2に動画画像符号化システム図を示す。

このシステムでは、動き補償予測誤差信号に対して Matching Pursuits 符号化を適用する。Matching Pursuits 符号化の際に得られた情報を用いて、各フレームで辞書学習を行い、辞書の更新情報を受信側に送信する。次のフレームの符号化時には、受信した情報により更新された辞書を用いて符号化を行う。したがって、本システムでは辞書更新情報が新たな付加伝送情報となる。

3.1 動的辞書設計アルゴリズム

次に本論文で提案する動的学習による辞書設計法について説明する。

動的辞書学習においては、各フレームに対して、2.3節で述べた辞書設計アルゴリズムを適用する。ただし、学習の際に用いる学習情報は、符号化したフレームのみの適用辞書パラメータと符号化対象となったブロックの画素値情報となる。また、学習における繰り返し処理は、各フレームで1回のみとする。

3.2 辞書更新情報と学習時間

提案法である動的学習辞書を用いる場合、新たな付加情報として辞書更新情報の伝送が必要となる。この辞書更新情報が大きい場合、再生品質が向上しても、全体としての符号化特性が低下する恐れがある。

そこで本節では、辞書更新のための情報量を抑えることについて検討する。まず、辞書更新のために伝送する情報は、波形自体ではなく、式(7)に示す Gabor 関数のパラメータとする。また、各パラメータの更新値についても、制限された範囲の離散値についてのみ評価することで情報量と処理時間を抑制する。

本論文では、学習によって更新するパラメータは、辞書パターンの構成に影響の大きい s と ξ とする。 ϕ については、従来より辞書パターンを空間的に重複して適用して

表1 更新された辞書パラメータの変化量
Renewed amounts of dictionary parameters.

変化量	Parameter	
	s	ξ
0	70.0%	94.6%
-1 ~ +1	96.3%	98.8%
-2 ~ +2	97.6%	99.6%
-3 ~ +3	99.3%	99.8%

表2 更新されたパラメータ数の割合
Ratio of renewed parameters in dictionary.

	Parameter	
	s	ξ
更新なし	74.6%	94.2%
更新あり	25.4%	5.8%

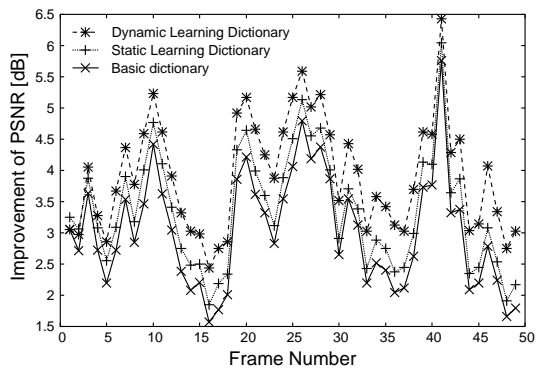
いることから、変更の効果が少ないものとして更新対象としない。また更新可能な値は、 s は 0~15 まで 1 間隔、 ξ は 0~7 まで 1 間隔の値とする。これらは今回用いる辞書ブロックのサイズ N が 16 画素であることから、設定した値である。

次にテストシーケンス carphone に対して、アトム数 400 で Matching Pursuits を行い、実際に動的学習辞書を用いた場合の、各フレームにおける辞書パラメータの変化量を調査した。結果を表1に示す。表1は、更新された変化量が一定範囲内となる辞書パラメータ数の全更新辞書パラメータに対する割合を示したものである。表1より、動的学習辞書におけるパラメータの変化量は、大部分が ± 1 の範囲に収まることが確認できる。これより、大幅なパラメータの更新が想定される最初のフレームを除いた、2フレーム目以降については、 s, ξ 共に $-1, 0, +1$ の3つに評価の値を限定することで情報量と処理時間を制限する。つまり、静的学習におけるパラメータ選択の際には、前述した更新可能な全てのパラメータ値について網羅的に式(9)による評価を行い、最も評価値が高いものにパラメータを更新することになるが、動的学習では、(現在のパラメータ値 ± 1) のパラメータ値についてのみ評価を行い、その中で最も評価値が高いパラメータに更新する。

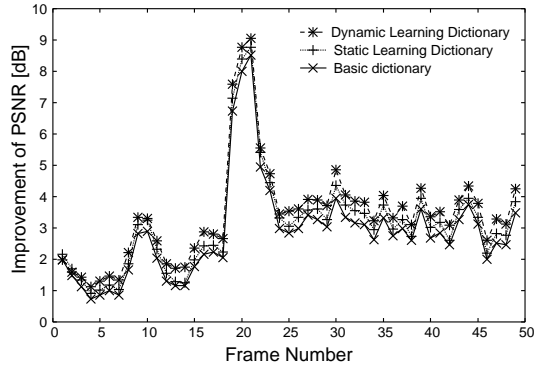
次に動的学習辞書において、各フレームで更新された辞書パターンの全体数に対する割合を表2に示す。表2より、各フレームにおいて更新された割合はパラメータ s が全体の $1/4$ 、 ξ が $1/20$ 程度である。

4. シミュレーション実験

提案法の性能評価のために計算機によるシミュレーション実験を行った。テストシーケンスには標準動画 mother & daughter, carphone(qcif, Y信号のみ, 50frames, 10fps)を用いた。シミュレーション実験では、図1に示す基本辞書、静的学習による辞書、提案法である動的学習による辞書について比較を行う。ここで静的辞書は、テストシーケンスを含まない10種類のシーケンス(boat, bus, bicycle, cheer leader, festival, football, girl, kiel, mobile&calendar,



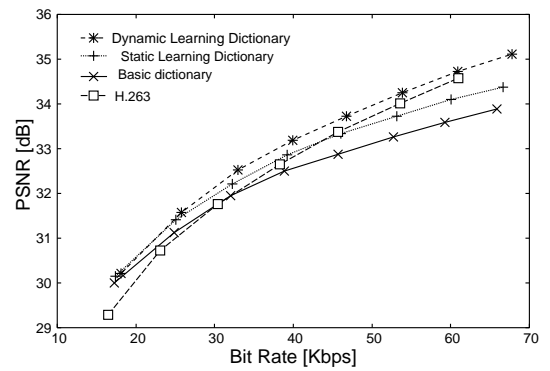
(a) carphone (400 atom)



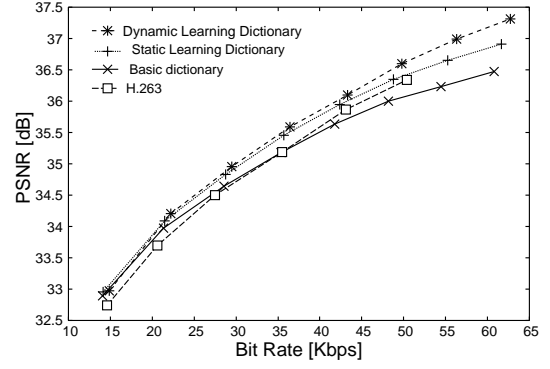
(b) mother & daughter (400 atom)

図 3 MP の各フレームにおける符号化特性比較

Improvement of coding performance of MP at each frame.



(a) carphone



(b) mother & daughter

図 4 符号化特性
Coding performance.

表 3 BD-PSNR による比較
Comparison by BD-PSNR.

シーケンス	BD-PSNR [dB]
carphone	0.408
mother & daughter	0.274

table tennis) から取得した学習データにより設計したものである。また、動的学習辞書における初期辞書は基本辞書を用いている。

まず、図 3 に各フレームにおける予測画像からの Matching Pursuits による PSNR の上昇量の比較を示す。図 3 より、ほぼすべてのフレームにおいて、動的学習辞書を用いた Matching Pursuits の品質が、基本辞書と静的学習辞書を上回るという結果が得られた。これは、動的学習により、フレーム毎に適した辞書への更新が実現されている効果であると考えられる。

次に、符号化特性の比較を図 4 に示す。DCT との比較として、同程度の符号化レートを対象とした H.263 の符号化特性も示す。H.263 については“TMN(H.263) coder ver.3.2”を用い、符号化レートを 10kbps から 60kbps 前後で一定間隔に設定して測定を行った。提案法においては、Matching Pursuits 部以外は H.263 と同様の手法とパラメータを用いている。ただし、先頭フレームは H.263 のイントラ符号化を適用した。

動的学習辞書を用いる場合には、辞書パラメータの更新情報を付加伝送しなければならない。しかしながら、その増加した情報量に対し、PSNR の改善量が大きいため、全体としての符号化特性が他の辞書を用いた場合より向上していることが確認できる。これらの結果より、動的学習辞書を用いることで Matching Pursuits の符号化特性の改善が実現した。

また、H.263 との比較評価として、提案法の H.263 に対する 16kbps から 64kbps における BD-PSNR(平均 PSNR 改善量)⁹⁾を表 3 に示す。表 3 より、提案法において、mother & daughter で約 0.27 dB, carphone で約 0.4 dB の品質改善が得られており、H.263 を上回る符号化特性が得られることが確認できた。H.263 に比べて、低ビットレートで Matching Pursuits の符号化特性が高い傾向にあるのは、DCT と異なり、画面全体ではなく近似効率が低い位置から逐次的に辞書による波形近似を行う方式であるためだと考えられる。

次に、提案法における情報量と処理時間の検証結果を示す。提案法では、動的学習により更新されるパラメータ s , ξ の変化量の組 (例えば, $+1, 0$ や $-1, +1$) に対して、テストシーケンスにおける各組の発生頻度からエントロピー符号化によって符号を割り当てている。実際に算出した 1 フレーム当たりの平均情報量は、全体で 6598 bit, そのうちの付加情報は 55 bit であり、全体の情報量の約 0.8% であった。したがって、動的辞書を用いることによる情報量のオーバーヘッドは少なく抑えられている。

また、提案法における辞書学習部などの各処理時間の全体に対する割合を表 4 に示す。表に示した割合は、アトム

表 4 処理時間 (200 アトム, 50 フレーム)
Processing times. (200 atoms, 50 frames)

処理	割合 [%]
MC	31.0
MP	66.7
辞書学習	2.0
その他	0.9

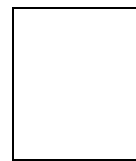
数 200 で carphone シーケンスを符号化した際の結果である。テストシーケンスの種類を変更しても、この割合にほとんど変化は見られなかった。表 4 より、学習における繰り返し処理を行わないこと、およびパラメータ評価値の制限により、学習のための処理時間は、符号化処理時間全体の 2% であり、非常に小さいことが確認できた。

5. む す び

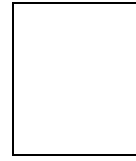
本論文では、Matching Pursuits 符号化特性の改善を目的として、学習的手法により各フレームの特性に適応した辞書へ更新する動的学習による辞書設計法を提案した。シミュレーションより、提案法を用いることで、複数シーケンスを用いた静的学習法による汎用辞書より、高い符号化特性を得ることができることが確認できた。今後は、階層型辞書等、高度な辞書構造と本提案法を組合せたさらなる特性改善を検討する予定である。

〔文 献〕

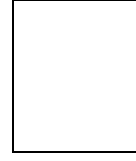
- 1) R. Neff, A. Zakhor: "Very Low Bit-Rate Video Coding Based on Matching Pursuits", IEEE Trans. Circuits & Syst. Video Technol., **7**, 1, pp.158-171 (1997)
- 2) R. Banham, C. Brailean: "A Selective Update Approach to Matching Pursuits Video Coding", IEEE Trans. Circuits & Syst. Video Technol., **7**, 1, pp.119-129 (1997)
- 3) S. Mallat, Z. Zhang: "Matching Pursuits With Time-Frequency Dictionaries", IEEE Trans. Signal Process., **41**, 12, pp.3397-3415 (1993)
- 4) R. Neff, A. Zakhor: "Matching Pursuits Video Coding. I. Dictionary Approximation", IEEE Trans. Circuits & Syst. Video Technol., **12**, 1, pp.13-26 (2002)
- 5) C.De. Vleeschouwer, B. Macq: "Subband Dictionaries for Low-Cost Matching Pursuits of Video Residues", IEEE Trans. Circuits & Syst. Video Technol., **9**, 7, pp.984-993 (1999)
- 6) 今村 幸祐, 小場 由雅, 橋本 秀雄: "動画画像符号化におけるサブバンド分割を用いた高速 Matching Pursuits アルゴリズム", 映情学誌, **60**, 7, pp 1059-1067 (2006)
- 7) P.S.-Saugeon, A.Zakhor: "Dictionary Design for Matching Pursuit and Application to Motion Compensated Video Coding", IEEE Trans. Circuits & Syst. Video Technol., **14**, 6, pp.880-886 (2004)
- 8) K.Imamura, S.Kako, H.Hashimoto: "Matching Pursuits using Slant Patterns and Its Dictionary Design" Proc. ISCAS 2005, pp.2080-2083 (2005)
- 9) G. Bjontegaard: "Calculation of average PSNR differences between RD-Curves", ITU-T Q.6/SG16 VCEG, VCEG-M33 (2001)



いまむら こうすけ
今村 幸祐 1995 年, 長崎大学工学部電気情報工学科卒業。1997 年, 同大大学院工学研究科修了。2000 年, 同大大学院海洋生産科学研究科修了。同年, 金沢大学工学部情報システム工学科助手。2005 年より, 同大大学院自然科学研究科講師。博士 (工学)。主として画像符号化に関する研究に従事。



たまき たのる
田巻 歩 2007 年, 金沢大学工学部情報システム工学科卒業。2009 年, 金沢大学大学院自然科学研究科博士前期課程修了見込。



はしもと ひろあ
橋本 秀雄 1968 年, 大阪大学工学部卒業。1970 年, 同大大学院修士課程修了。1975 年, 同大大学院博士課程修了。同年, 電電公社 (現 NTT) 電気通信研究所入所。以来, 画像符号化方式, 画像通信方式の研究開発に従事。現在, 金沢大学大学院自然科学研究科教授。工学博士。正会員。